COMMENTS OF BELLSOUTH

CC DOCKET NO. 01-318

JANUARY 22, 2002

ATTACHMENT 2, PART 4

Appendix C
Balancing the Type I and Type II Error Probabilities of the Truncated Z Test Statistic

This appendix describes a the methodology for balancing the error probabilities when the Truncated Z statistic, described in Appendix A, is used for performance measure parity testing. There are four key elements of the statistical testing process:

1. the null hypothesis, H_0, that parity exists between ILEC and CLEC services
2. the alternative hypothesis, H_a, that the ILEC is giving better service to its own customers
3. the Truncated Z test statistic, Z^T, and
4. a critical value, c

The decision rule[1] is

- If $Z^T < c$ then accept H_a.
- If $Z^T \geq c$ then accept H_0.

There are two types of error possible when using such a decision rule:

Type I Error: Deciding favoritism exists when there is, in fact, no favoritism.
Type II Error: Deciding parity exists when there is, in fact, favoritism.

The probab lities of each type of each are:

Type I Error: $\alpha = P(Z^T < c \mid H_0)$.
Type II Error: $\beta = P(Z^T \geq c \mid H_a)$.

In what follows, we show how to find a balancing critical value, c_B, so that $\alpha = \beta$.

General Methodology

The general form of the test statistic that is being used is

[1] This decision rule assumes that a negative test statistic indicates poor service for the CLEC customer. If the opposite is true, then reverse the decision rule.

$$z_0 = \frac{\hat{T} - E(\hat{T}|H_0)}{SE(\hat{T}|H_0)}, \tag{C.1}$$

where

$\hat{T}$ is an estimator that is (approximately) normally distributed,

$E(\hat{T} \mid H_0)$ is the expected value (mean) of $\hat{T}$ under the null hypothesis, and

$SE(\hat{T} \mid H_0)$ is the standard error of $\hat{T}$ under the null hypothesis.

Thus, under the null hypothesis, z_0 follows a standard normal distribution. However, this is not true under the alternative hypothesis. In this case,

$$z_a = \frac{\hat{T} - E(\hat{T}|H_a)}{SE(\hat{T}|H_a)}$$

has a standard normal distribution. Here

$E(\hat{T} \mid H_a)$ is the expected value (mean) of $\hat{T}$ under the alternative hypothesis, and

$SE(\hat{T} \mid H_a)$ is the standard error of $\hat{T}$ under the alternative hypothesis.

Notice that

$$\begin{aligned} \beta &= P(z_0 > c \mid H_a) \\ &= P\left(z_a > \frac{cSE(\hat{T} \mid H_0) + E(\hat{T} \mid H_0) - E(\hat{T} \mid H_a)}{SE(\hat{T} \mid H_a)} \right) \end{aligned} \tag{C.2}$$

and recall that for a standard normal random variable z and a constant b, $P(z < b) = P(z > -b)$. Thus,

$$\alpha = P(z_0 < c) = P(z_0 > -c) \tag{C.3}$$

Since we want $\alpha = \beta$, the right hand sides of (C.2) and (C.3) represent the same area under the standard normal density. Therefore, it must be the case that

$$-c = \frac{cSE(\hat{T} \mid H_0) + E(\hat{T} \mid H_0) - E(\hat{T} \mid H_a)}{SE(\hat{T} \mid H_a)}.$$

Solving this for c gives the general formula for a balancing critical value:

$$c_B = \frac{E(\hat{T}\mid H_a) - E(\hat{T}\mid H_0)}{SE(\hat{T}\mid H_a) + SE(\hat{T}\mid H_0)} \qquad (C.4)$$

The Balancing Critical Value of the Truncated Z

In Appendix A, the Truncated Z statistic is defined as

$$Z^T = \frac{\sum_j W_j Z_j^* - \sum_j W_j E(Z_j^*\mid H_0)}{\sqrt{\sum_j W_j^2 Var(Z_j^*\mid H_0)}}$$

In terms of equation (C.1) we have

$$\hat{T} = \sum_j W_j Z_j^*$$

$$E(\hat{T}\mid H_0) = \sum_j W_j E(Z_j^*\mid H_0)$$

$$SE(\hat{T}\mid H_0) = \sqrt{\sum_j W_j^2 Var(Z_j^*\mid H_0)}$$

To compute the balancing critical value (C.4), we also need $E(\hat{T}\mid H_a)$ and $SE(\hat{T}\mid H_a)$. These values are determined by

$$E(\hat{T}\mid H_a) = \sum_j W_j E(Z_j^*\mid H_a), \text{ and}$$

$$SE(\hat{T}\mid H_a) = \sqrt{\sum_j W_j^2 \, var(Z_j^*\mid H_a)}.$$

In which case equation (C.4) gives

$$c_B = \frac{\sum_j W_j E(Z_j^*\mid H_a) - \sum_j W_j E(Z_j^*\mid H_0)}{\sqrt{\sum_j W_j^2 \, var(Z_j^*\mid H_a)} + \sqrt{\sum_j W_j^2 \, var(Z_j^*\mid H_0)}}. \qquad (C.5)$$

Thus, we need to determine how to calculate $E(Z_j^*\mid H_0)$, $Var(Z_j^*\mid H_0)$, $E(Z_j^*\mid H_a)$, and $Var(Z_j^*\mid H_a)$.

If Z_j has a normal distribution with mean μ and standard error σ, then the mean of the distribution truncated at 0 is

$$M(\mu,\sigma) = \int_{-\infty}^{0} \frac{x}{\sqrt{2\pi}\sigma} \exp(-\tfrac{1}{2}\left(\tfrac{x-\mu}{\sigma}\right)^2)dx\,,$$

and the variance is

$$V(\mu,\sigma) = \int_{-\infty}^{0} \frac{x^2}{\sqrt{2\pi}\sigma} \exp(-\tfrac{1}{2}\left(\tfrac{x-\mu}{\sigma}\right)^2)dx - M(\mu,\sigma)^2$$

It can be shown that

$$M(\mu,\sigma) = \mu\Phi(\tfrac{-\mu}{\sigma}) - \sigma\phi(\tfrac{-\mu}{\sigma})$$

and

$$V(\mu,\sigma) = (\mu^2+\sigma^2)\Phi(\tfrac{-\mu}{\sigma}) - \mu\sigma\phi(\tfrac{-\mu}{\sigma}) - M(\mu,\sigma)^2$$

where $\Phi(\cdot)$ is the cumulative standard normal distribution function, and $\phi(\cdot)$ is the standard normal density function.

The cell test statistic, Z_j, is constructed so that it has mean 0 and standard deviation 1 under the null hypothesis. Thus,

$$E(Z_j^*|H_0) = M(0,1) = -\frac{1}{\sqrt{2\pi}}\,,\text{ and}$$

$$\text{var}(Z_j^*|H_0) = V(0,1) = \frac{1}{2} - \frac{1}{2\pi}\,.$$

The mean and standard error of Z_j under the alternative hypothesis depends on the type of measure and the form of the alternative. These are discussed below. For now, denote the mean and standard error of Z_j under the alternative by m_j and se_j respectively. Thus,

$$E(Z_j^*|H_a) = M(m_j,se_j)\,,\text{ and}$$

$$SE(Z_j^*|H_a) = V(m_j,se_j)\,.$$

Using the above notation, and equation (C.5), we get the formula for the balancing critical of Z^T.

$$c_B = \frac{\sum_j W_j M(m_j, se_j) - \sum_j W_j \frac{-1}{\sqrt{2\pi}}}{\sqrt{\sum_j W_j^2 V(m_j, se_j)} + \sqrt{\sum_j W_j^2 \left(\frac{1}{2} - \frac{1}{2\pi}\right)}}. \qquad (C.6)$$

This formu a assumes that Z_j is approximately normally distributed within cell j. When the cell sample sizes, n_{1j} and n_{2j}, are small this may not be true. It is possible to determine the cell mean and variance under the null hypothesis when the cell sample sizes are small. It is much more difficult to determine these values under the alternative hypothesis. Since the cell weight, W_j will also be small (see Appendix A) for a cell with small volume, the cell mean and variance will not contribute much to the weighted sum. Therefore, formula (C.6) provides a reasonable approximation to the balancing critical value.

Alternative Hypotheses

Mean Measure

For mean measures, one is concerned with two parameters in each cell, namely, the mean and variance. A possible lack of parity may be due to a difference in cell means, and/or a difference in cell variances. One possible set of hypotheses that capture this notion, and take into account the assumption that transaction are identically distributed within cells is:

$$H_0: \mu_{1j} = \mu_{2j}, \sigma_{1j}^2 = \sigma_{2j}^2$$

$$H_a: \mu_{2j} = \mu_{1j} + \delta_j \cdot \sigma_{1j}, \sigma_{2j}^2 = \lambda_j \cdot \sigma_{1j}^2 \qquad \delta_j > 0, \lambda_j \geq 1 \text{ and } j = 1, \ldots, L.$$

Under this form of alternative hypothesis, the cell test statistic Z_j has mean and standard error given by

$$m_j = \frac{-\delta_j}{\sqrt{\frac{1}{n_{1j}} + \frac{1}{n_{2j}}}}, \text{ and}$$

$$se_j = \sqrt{\frac{\lambda_j n_{1j} + n_{2j}}{n_{1j} + n_{2j}}}$$

Proportion Measure

For a proportion measure there is only one parameter of interest in each cell, the proportion of transaction possessing an attribute of interest. A possible lack of parity may be due to a difference in cell proportions. A set of hypotheses that take into account the assumption that transaction are identically distributed within cells while allowing for an analytically tractable solution is:

$$H_0: \frac{p_{2j}(1-p_{1j})}{(1-p_{2j})p_{1j}} = 1$$

$$H_a: \frac{p_{2j}(1-p_{1j})}{(1-p_{2j})p_{1j}} = \psi_j \qquad \psi_j > 1 \text{ and } j = 1,\ldots,L.$$

These hypotheses are based on the "odds ratio." If the transaction attribute of interest is a missed trouble repair, then an interpretation of the alternative hypothesis is that a CLEC trouble is ψ_j times more likely to be missed than an ILEC trouble.

Under this form of alternative hypothesis, the within cell asymptotic mean and variance of a_{1j} are given by[2]

$$E(a_{1j}) = n_j \pi_j^{(1)}$$

$$\text{var}(a_{1j}) = \frac{n_j}{\frac{1}{\pi_j^{(1)}} + \frac{1}{\pi_j^{(2)}} + \frac{1}{\pi_j^{(3)}} + \frac{1}{\pi_j^{(4)}}} \tag{C.7}$$

where

$$\pi_j^{(1)} = f_j^{(1)}\left(n_j^2 + f_j^{(2)} + f_j^{(3)} - f_j^{(4)}\right)$$

$$\pi_j^{(2)} = f_j^{(1)}\left(-n_j^2 - f_j^{(2)} + f_j^{(3)} + f_j^{(4)}\right)$$

$$\pi_j^{(3)} = f_j^{(1)}\left(-n_j^2 + f_j^{(2)} - f_j^{(3)} + f_j^{(4)}\right)$$

$$\pi_j^{(4)} = f_j^{(1)}\left(n_j^2\left(\frac{2}{\psi_j} - 1\right) - f_j^{(2)} - f_j^{(3)} - f_j^{(4)}\right)$$

$$f_j^{(1)} = \frac{1}{2n_j^2\left(\frac{1}{\psi_j} - 1\right)}$$

$$f_j^{(2)} = n_j n_{1j}\left(\frac{1}{\psi_j} - 1\right)$$

$$f_j^{(3)} = n_j a_j\left(\frac{1}{\psi_j} - 1\right)$$

$$f_j^{(4)} = \sqrt{n_j^2\left[4n_{1j}\left(n_j - a_j\right)\left(\frac{1}{\psi_j} - 1\right) + \left(n_j + \left(a_j - n_{1j}\right)\left(\frac{1}{\psi_j} - 1\right)\right)^2\right]}$$

[2] Stevens, W. L. (1951) Mean and Variance of an entry in a Contingency Table. *Biometrica*, **38**, 468-470.

Recall that the cell test statistic is given by

$$Z_j = \frac{n_j a_{1j} - n_{1j} a_j}{\sqrt{\dfrac{n_{1j} n_{2j} a_j (n_j - a_j)}{n_j - 1}}}.$$

Using the equations in (C.7), we see that Z_j has mean and standard error given by

$$m_j = \frac{n_j^2 \pi_j^{(1)} - n_{1j} a_j}{\sqrt{\dfrac{n_{1j} n_{2j} a_j (n_j - a_j)}{n_j - 1}}}, \text{ and}$$

$$se_j = \sqrt{\frac{n_j^3 (n_j - 1)}{n_{1j} n_{2j} a_j (n_j - a_j)\left(\frac{1}{\pi_j^{(1)}} + \frac{1}{\pi_j^{(2)}} + \frac{1}{\pi_j^{(3)}} + \frac{1}{\pi_j^{(4)}}\right)}}.$$

Rate Measure

A rate measure also has only one parameter of interest in each cell, the rate at which a phenomenon is observed relative to a base unit, e.g. the number of troubles per available line. A possible lack of parity may be due to a difference in cell rates. A set of hypotheses that take into account the assumption that transaction are identically distributed within cells is:

$$H_0: r_{1j} = r_{2j}$$

$$H_a: r_{2j} = \varepsilon_j r_{1j} \qquad \varepsilon_j > 1 \text{ and } j = 1,\ldots,L.$$

Given the total number of ILEC and CLEC transactions in a cell, n_j, and the number of base elements, b_{1j} and b_{2j}, the number of ILEC transaction, n_{1j}, has a binomial distribution from n_j trials and a probability of

$$q_j^* = \frac{r_{1j} b_{1j}}{r_{1j} b_{1j} + r_{2j} b_{2j}}.$$

Therefore, the mean and variance of n_{1j}, are given by

$$\begin{aligned} E(n_{1j}) &= n_j q_j^* \\ \operatorname{var}(n_{1j}) &= n_j q_j^* (1 - q_j^*) \end{aligned} \qquad \text{(C.8)}$$

Under the null hypothesis

$$q_j^* = q_j = \frac{b_{1j}}{b_j},$$

but under the alternative hypothesis

$$q_j^* = q_j^a = \frac{b_{1j}}{b_{1j} + \varepsilon_j b_{2j}}. \tag{C.9}$$

Recall that the cell test statistic is given by

$$Z_j = \frac{n_{1j} - n_j q_j}{\sqrt{n_j q_j (1 - q_j)}}.$$

Using (C.8) and (C.9), we see that Z_j has mean and standard error given by

$$m_j = \frac{n_j \left(q_j^a - q_j\right)}{\sqrt{n_j q_j (1 - q_j)}} = (1 - \varepsilon_j)\sqrt{\frac{n_j b_{1j} b_{2j}}{b_{1j} + \varepsilon_j b_{2j}}}, \text{ and}$$

$$se_j = \sqrt{\frac{q_j^a (1 - q_j^a)}{q_j (1 - q_j)}} = \sqrt{\varepsilon_j}\frac{b_j}{b_{1j} + \varepsilon_j b_{2j}}.$$

Ratio Measure

As with mean measures, one is concerned with two parameters in each cell, the mean and variance, when testing for parity of ratio measures. As long as sample sizes are large, as in the case of billing accuracy, the same method for finding m_j and se_j that is used for mean measures can be used for ratio measures.

Determining the Parameters of the Alternative Hypothesis

In this appendix we have indexed the alternative hypothesis of mean measures by two sets of parameters, λ_j and δ_j. Proportion and rate measures have been indexed by one set of parameters each, ψ_j and ε_j respectively. A major difficulty with this approach is that more than one alternative will be of interest; for example we may consider one alternative in which all the δ_j are set to a common non-zero value, and another set of alternatives in each of which just one δ_j is non-zero, while all the rest are zero. There are very many other possibilities. Each possibility leads to a single value for the balancing critical value; and each possible critical value corresponds to many sets of alternative hypotheses, for each of which it constitutes the correct balancing value.

The formulas we have presented can be used to evaluate the impact of different choices of the overall critical value. For each putative choice, we can evaluate the set of alternatives for which this is the correct balancing value. While statistical science can be used to evaluate the impact of different choices of these parameters, there is not much that an appeal to statistical principles can offer in directing specific choices. Specific choices are best left to telephony experts. Still, it is possible to comment on some aspects of these choices:

- Parameter Choices for λ_j. The set of parameters λ_j index alternatives to the null hypothesis that arise because there might be greater unpredictability or variability in the delivery of service to a CLEC customer over that which would be achieved for an otherwise comparable ILEC customer. While concerns about differences in the variability of service are important, it turns out that the truncated Z testing which is being recommended here is relatively insensitive to all but very large values of the λ_j. Put another way, reasonable differences in the values chosen here could make very little difference in the balancing points chosen.

- Parameter Choices for δ_j. The set of parameters δ_j are much more important in the choice of the balancing point than was true for the λ_j. The reason for this is that they directly index differences in average service. The truncated Z test is very sensitive to any such differences; hence, even small disagreements among experts in the choice of the δ_j could be very important. Sample size matters here too. For example, setting all the δ_j to a single value – $\delta_j = \delta$ might be fine for tests across individual CLECs where currently in Louisiana the CLEC customer bases are not too different. Using the same value of δ for the overall state testing does not seem sensible. At the state level we are aggregating over CLECs, so using the same δ as for an individual CLEC would be saying that a "meaningful" degree of disparity is one where the violation is the same (δ) for each CLEC. But the detection of disparity for any component CLEC is important, so the relevant "overall" δ should be smaller.

- Parameter Choices for ψ_j or ε_j. The set of parameters ψ_j or ε_j are also important in the choice of the balancing point for tests of their respective measures. The reason for this is that they directly index increases in the proportion or rate of service performance. The truncated Z test is sensitive to such increases; but not as sensitive as the case of δ for mean measures. Sample size matters here too. As with mean measures, using the same value of ψ or ε for the overall state testing does not seem sensible.

The three parameters are related however. If a decision is made on the value of δ, it is possible to determine equivalent values of ψ and ε. The following equations, in conjunction with the definitions of ψ and ε, show the relationship with delta.

$$\delta = 2 \cdot \arcsin(\sqrt{\hat{p}_2}) - 2 \cdot \arcsin(\sqrt{\hat{p}_1})$$
$$\delta = 2\sqrt{\hat{r}_2} - 2\sqrt{\hat{r}_1}$$

The bottom line here is that beyond a few general considerations, like those given above, a principled approach to the choice of the alternative hypotheses to guard against must come from elsewhere.

Appendix D: Examples of Statistical Reports

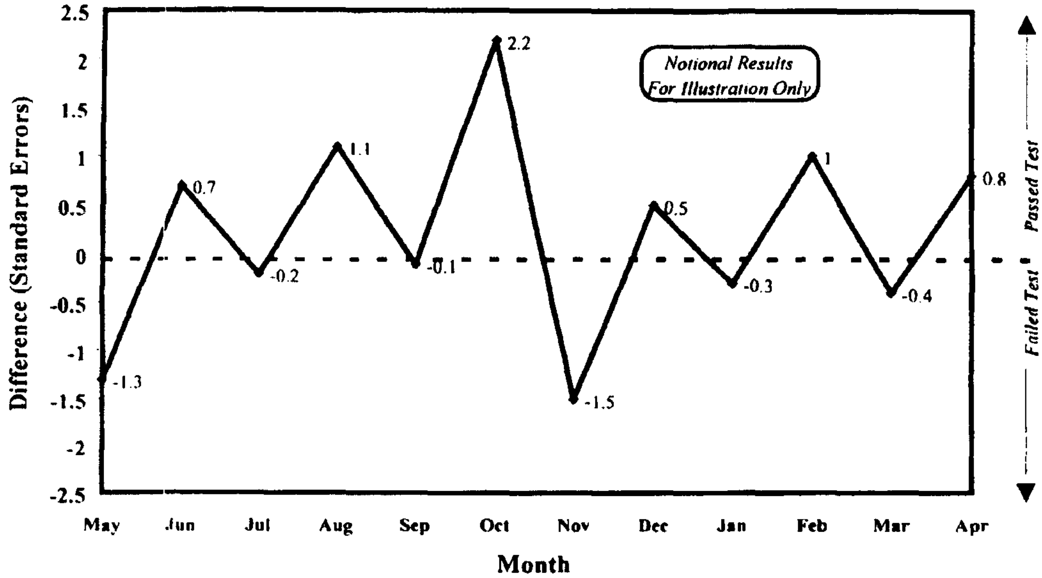
The general structure for reporting statistical results in a production environment will be the same for the different measures and we suggest that it consist of at least three components. For each measure present, (1) the monthly test statistics over a period of time, (2) the results for the current month, with summary statistics, test statistics, and descriptive graphs, and (3) a summary of any adjustments to the data made in the process of running the tests, including a description of how many records were excluded from analysis and the reason for the exclusion (i.e., excluded due to business rules, or due to statistical/methodological rules pertaining to the measure). The last component is important to assure that the reported results can be audited.

Selected components of the reporting structure are illustrated in the samples that follow. An outline of the report is shown below. Monthly results will be presented for each level of aggregation required.

I. Test Statistics Over Time
II. Monthly Results
   A. Summary Statistics
   B. Test Statistics
   C. Descriptive Graphs (Frequency Distributions, etc.)
III. Adjustments to Data
   A. Records Excluded Due to Business Rules
   B. Records Excluded Due to Statistical Rules

Test Statistic Over Time. The first component of the reporting structure is an illustration of the trend of the particular performance measure over time together with a tabular summary of results for the current month. We will show at a glance whether the tests consistently return non-statistically significant results; consistently indicate disparity (be that in favor of BellSouth or in favor of the CLECs); or vary month by month in their results. An example of this component follows.

Notional Performance Measure
Through April XXXX
Differences Between Test Statistic and Balancing Critical Value

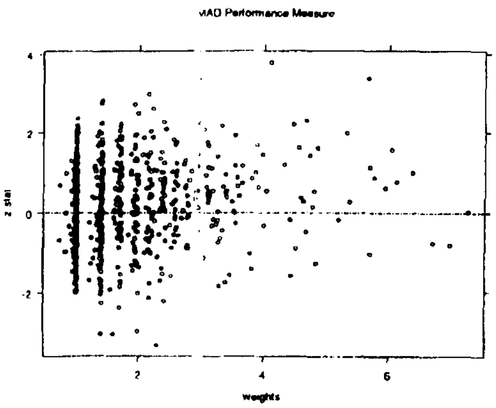


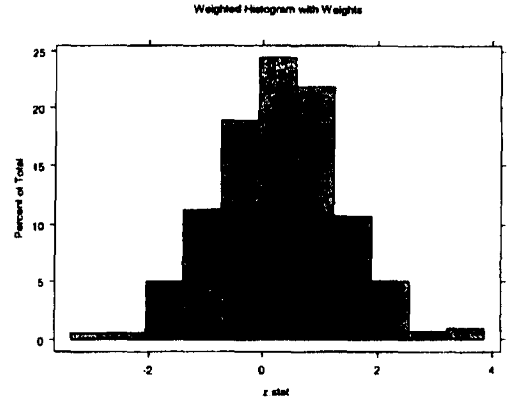
Result for Current Month	
Test Statistic	-0.410
Balancing Critical Value	-1.210
Difference	0.800

Monthly Results. The most important component of the reporting structure is the part which presents results of the monthly statistical tests on the given performance measure. The essential aspects included in this component are the summary statistics; the test statistics and results; and descriptive graphs of the results.

It is important to present basic summary statistics to complete the comparison between BellSouth and the CLECs. At a minimum, these statistics will include the means, standard deviations, and population sizes. In addition to basic descriptive statistics, we also present the test statistic results. Examples of ways we have presented these statistics in the past can be found in BellSouth's February 25, 1999 filing before the Louisiana Public Service Commission.

Finally, the results will be presented in graphical format. Below is an example of how to graphically present the data behind the Truncated Z statistic. One graph shows a plot of cell Z score versus cell weights. The other is a histogram of the weighted cell Z scores.





Adjustments to Data. The third important component of the reporting structure is information on any adjustments performed on the data. This information is essential in order that the results may be verified and audited. The most prevalent examples of such modifications would be removal of observations and weighting of the data.

Records can be removed from analysis for both business reasons (these will likely be taken into account in the PMAP system) and for statistical reasons. All of the performance measures exclude certain records based on business rules underlying each measure's particular definitions and methodologies. The number of records excluded for each rule will be summarized. In addition, some of the measures will have observations excluded for statistical reasons, particularly in the case of "mean measures" (OCI and MAD); these exclusions will be summarized as well. The tables below show examples of the current method for summarizing this information:

April XXXX
Peromance Measure Filtering Information

This table displays information about the size of the database files and the cases that were removed from the analysis.

		1999	
Unfiltered Total	**28,691**	**Unfiltered Total**	**453,107**
Records Removed for Business Reasons *(e.g. not N, T, C, or P orders, not resale and not UNE)*	**7,242**	**Records Removed for Business Reasons** *(e.g. not N, T, C, or P orders, not retail)*	**78,613**
Total Reported on Web Report	21,449	Total Reported on Web Report	374,494
Additional Records Removed for Business Reasons	**876**	**Additional Records Removed for Business Reasons**	**7,429**
Missing Appointment code is 'S'	844	Missing Appointment code is 'S'	7,172
General Class Service = 'O'	0	General Class Service = 'O'	279
UNE Cases	102		
Records Removed for Statistical Reasons		**Records Removed for Statistical Reasons**	
Extreme Values Removed	**9**	**Extreme Values Removed**	**652**
No Matching Classification Removals	**47**	**No Matching Classification Removals**	**21,974**
FILTERED TOTAL	20,517	FILTERED TOTAL	344,439

Appendix E. Trimming Outliers for Mean Measures

The arithmetic average is extremely sensitive to outliers; a single large value, possibly an erroneous value, can significantly distort the mean value. And by inflating the error variance, this also affects conclusions in the test of hypotheses. Extreme data values may be correct, but since they are rare measurements, they may be considered to be statistical outliers. Or they may be values that should not be in the analysis data set because of errors in the measurement or in selecting the data.

At this time, only two mean measures have been analyzed: Order Completion Interval and Maintenance Average Duration. Maintenance Average Duration data are truncated at 240 hours and therefore this measure was not trimmed further. For Order Completion Interval, the underlying distribution of the observations is clearly not normal, but rather skewed with a very long upper-tail.

A useful technique, coming from the field of robust statistical analysis, is to trim a very small proportion from the tails of the distribution before calculating the means. The resulting mean is referred to as a trimmed mean. Trimming is beneficial in that it speeds the convergence of the distribution of the means to a normal distribution. Only extreme values are trimmed, and in many cases the data being trimmed are, in fact, data that might not be used in the analysis on other grounds.

In the first analysis of the verified Order Completion Interval-Provisioning measure, after removing data that were clearly in error or were not applicable, we looked at the cases that represented the largest 0.01% of the BST distribution. In the August data, this corresponded to orders with completion intervals greater than 99 days. All of these were BellSouth orders. In examining the largest 11 individual examples that would be removed from analysis, we found that only 1 of the 11 cases was a valid case where the completion interval was unusually large. The other 10 cases were examples of cases that should not have been included in the analysis. This indicates that at least in preliminary analysis, it is both beneficial to examine the extreme outliers and reasonable to remove them.

A very slight trimming is needed in order to put the central limit theorem argument on firm ground. But finding a robust rule that can be used in a production setting is difficult. Also, any trimming rule should be fully explained and any observations that are trimmed from the data must be fully documented.

When it is determined that a measure should be trimmed, a trimming rule that is easy to implement in a production setting is:

> **Trim the ILEC observations to the largest CLEC value from all CLEC observations in the month under consideration.**

That is, no CLEC values are removed; all ILEC observations greater than the largest CLEC observation are trimmed.

While this method is simple, it does allow for extreme CLEC observations to be part of the analysis. For instance, suppose that the amount of time to complete an order was less than 40 days for all CLEC orders except one. Let's say that this extreme order took 100 days to complete. The trimming rule says that all ILEC orders above 100 days should be trimmed, but a closer look at the data might suggest trimming at 40 days instead.

Since we are operating in a production mode system, it is not possible to explore the data before the trimming takes place. Other automatic trimming rules present other problems, so our solution is to use the simple trimming rule above, and have the system automatically produce a trimming report that can be examined at a later point in time.

The trimming report should include:

- The value of the trim point.
- Summary statistics and graphics of the ILEC observations that were trimmed.
- A listing of the trimmed ILEC transaction for a random sample of 10 trimmed transactions. This listing should not disclose sensitive information.
- A listing of the 10 most extreme CLEC transactions. This listing should not disclose sensitive information.
- The number of ILEC and CLEC observations above some fixed point, so that changes in the upper tail can be better tracked over time.

The trimming report should be part of the overall report discussed in Appendix D. Examples of tables contained within the trimming report are shown below.

April XXXX
Performance Measure Extreme Values

CLEC		BST	
Cutoff	26	Cutoff	26
# of Records	20,573	# of Records	367,065
10 Largest		Extreme Values	652
Minimum	19	Minimum	27
Median	23	Median	32
Maximum	26	Maximum	283
Subtotal	20,573	Subtotal	366,413

April XXXX
Performance Measure Weighting Report

CLEC		BST	
# of Records	20,573	# of Records	366,413
No Matching BST		No Matching CLEC	
Classification (1)	47	Classification (2)	21,974
Subtotal	20,526	Subtotal	344,439

April XXXX
Peroramance Measure Filtering Information

This table displays information about the size of the database files and the cases that were removed from the analysis.

			1999	
Unfiltered Total	**28,691**		**Unfiltered Total**	**453,107**
Records Removed for Business Reasons *(e.g. not N, T, C, or P orders, not resale and not UNE)*	**7,242**		**Records Removed for Business Reasons** *(e.g. not N, T, C, or P orders, not retail)*	**78,613**
Total Reported on Web Report	**21,449**		**Total Reported on Web Report**	**374,494**
Additional Records Removed for Business Reasons	**876**		**Additional Records Removed for Business Reasons**	**7,429**
Missing Appointment code is 'S'	844		Missing Appointment code is 'S'	7,172
General Class Service = 'O'	0		General Class Service = 'O'	279
UNE Cases	102			
Records Removed for Statistical Reasons			**Records Removed for Statistical Reasons**	
Extreme Values Removed	**0**		**Extreme Values Removed**	**652**
No Matching Classification Removals	**47**		**No Matching Classification Removals**	**21,974**
FILTERED TOTAL	**20,526**		**FILTERED TOTAL**	**344,439**

CLEC Extreme Values

Wire Center	Time	Dispatch	Residence	Circuits	Order Type	Order Interval
NWORLAMA	1	1	3	1	N	61
OPLSLATL	1	2	1	1	C	53
NWORLAMA	2	1	3	1	N	44
NWORLAMA	1	1	3	1	N	39
BTRGLAWN	1	1	2	1	C	38
LKCHLADT	1	1	1	1	T	37
NWORLAMA	1	1	3	1	N	32
NWORLAMA	2	1	3	1	N	32
SHPTLACL	1	1	2	1	N	28

Frequency of Extreme Values Removed from BST file (Top 10)

Wire Center	Time	Dispatch	Residence	Circuits	Order Type	Frequency
NWORLAMA	1	1	3	1	N	55
NWORLAMA	2	1	3	1	N	25
BTRGLASB	2	1	3	1	C	23
NWORLAMC	2	1	3	1	C	23
NWORLAMC	1	1	3	1	C	22
NWORLAMA	2	1	3	1	C	18
NWORLAMA	1	1	3	1	C	17
BTRGLASB	1	1	3	1	C	16
LFYTLAMA	1	1	3	1	C	15
NWORLAMA	2	2	3	1	C	14